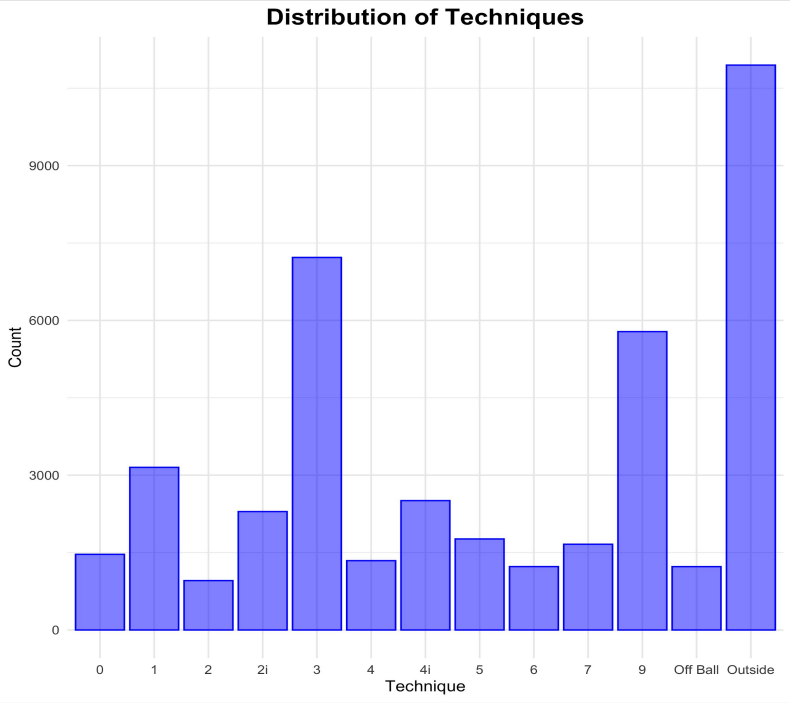
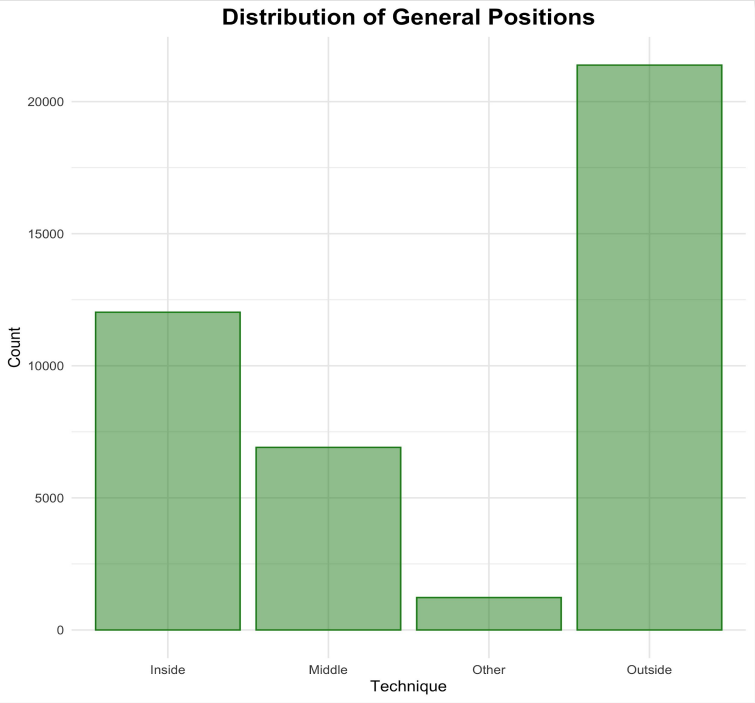
Evaluating the NFL's Trenches

By: Calvin Smith
Email: calvinsmith625@gmail.com

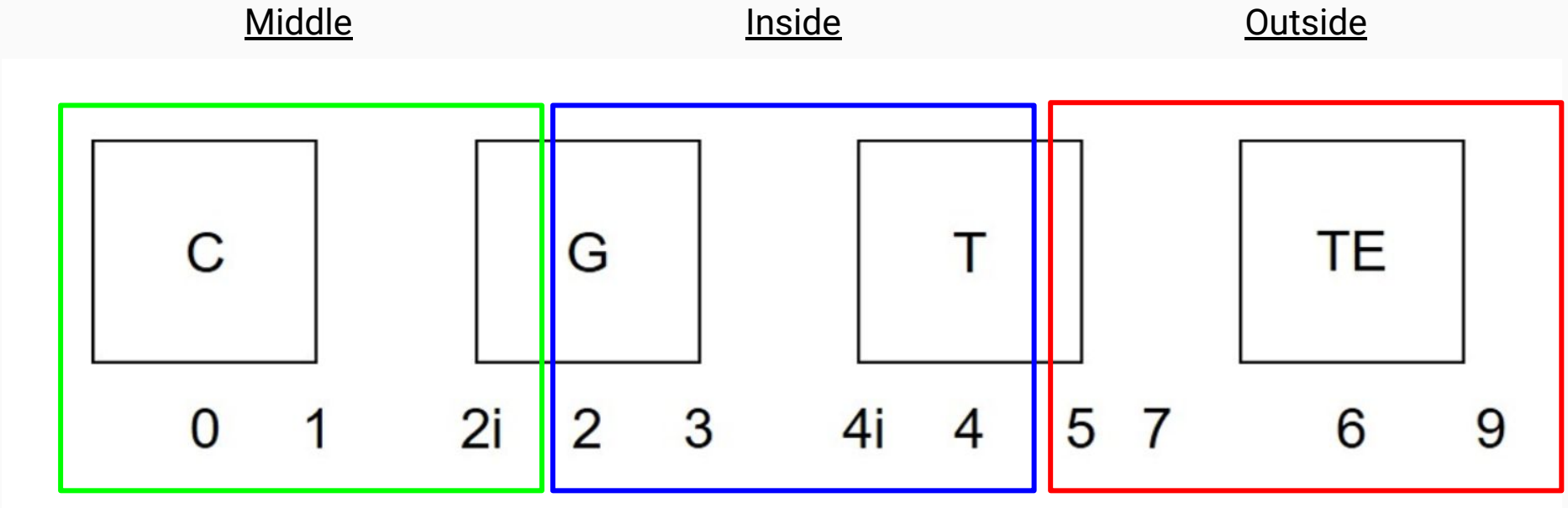
- **When evaluating the most important position along the defensive line, the first question is:**
  - How will we categorize the positions?
- **Other building block questions are:**
  - How granular should we get with the positional definitions?
  - How do we not conflate importance with talent?
- **The first step is to group players into more general alignments on the field. This will help find signal in where to look for positional value. These groupings may be different for the final answer.**

What are the most common D-Line positions?





How were the general alignments defined?



Now that we have a more general grouping to find signal, we can begin to analyze.

It's important to evaluate run and pass plays separately, so run plays were separated from passes. Let's start with pass plays, given those are more common in the data (~ 1.3 passes per rush) and are more costly from a defensive perspective. By EPA standards, the ability to affect the pass is more important than affecting the run.

In order to determine the most important position for slowing down a passing attack we'll look at pressures as our primary way of measuring how much a defensive lineman affects completions and pass EPA.

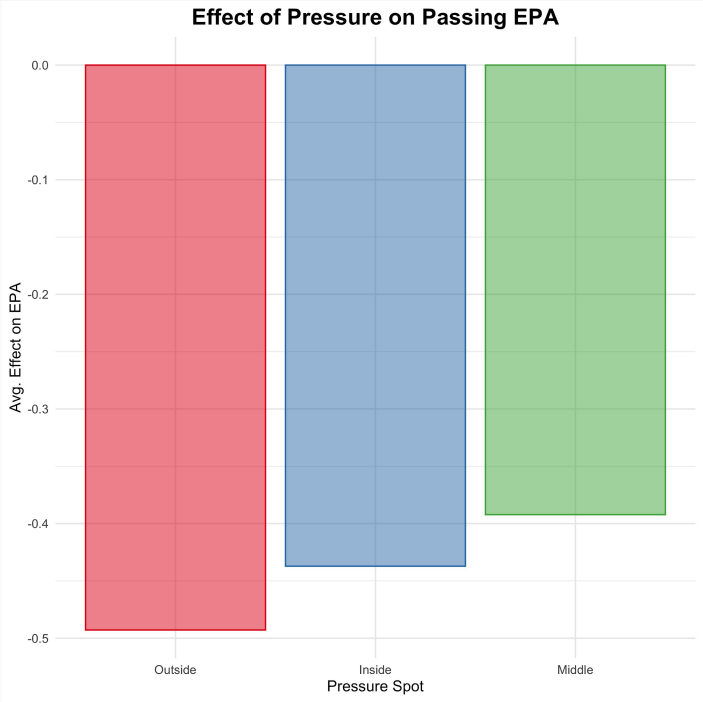
We turn to simple regression to check the observed influence of pressure on passing EPA. To see if pressure affects completions we'll use logistic regression to model the observed probability of completing passes with or without pressure on the QB.

How does pressure from generalized positions affect EPA?

- The regression output on the right reflects the effect of Pressure from the Middle, Inside, Outside, and No Pressure on passing EPA (spikes omitted).
- It's clear that having no pressure is critical for an offense to succeed, but pressure from the outside zone distinguishes itself as highly statistically significant opposed to inside and middle pressure.
    - This may be due to the existing emphasis on outside pass rushing, as well as a general speed discrepancy between players on the edge versus on the inside/middle.
- On average, outside defensive lineman's pressures affect opposing offense's passing EPA more than their counterparts along the line.

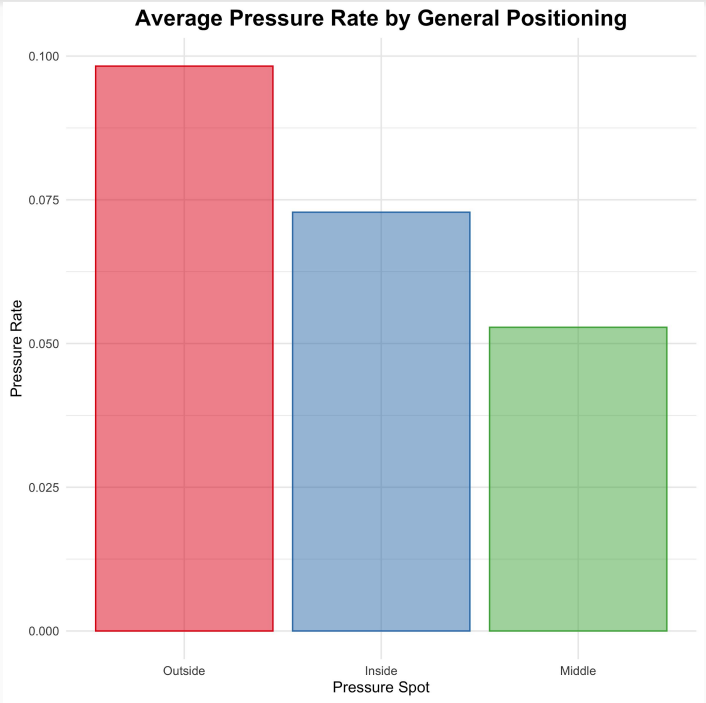
```
Call:
lm(formula = EPA ~ PrsByMiddle + PrsByInside + PrsByOutside +
    NoPrs, data = passes)

Residuals:
     Min       1Q   Median       3Q      Max
-10.7440  -0.7598  -0.1335   0.7813   7.1779

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.38374    0.01302 -29.468  < 2e-16 ***
PrsByMiddle  -0.01005    0.07492  -0.134 0.893242
PrsByInside  -0.05344    0.04937  -1.082 0.279131
PrsByOutside -0.11046    0.03338  -3.309 0.000938 ***
NoPrs         0.59139    0.01564  37.823  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.41 on 41543 degrees of freedom
Multiple R-squared:  0.04174,   Adjusted R-squared:  0.04165
F-statistic: 452.4 on 4 and 41543 DF,  p-value: < 2.2e-16
```

When an outside lineman pressures there is a distinguishable effect on the play's EPA.



Not only are outside pressures better for the defense, but occur more frequently.

- Here we observe that outside linemen generate pressure more often than the rest of the defensive line.
- So, the area that produces the greatest effect on EPA also creates pressure more often.
    - If you're looking to create the best pass rush, you're likely best suited to build that from the outside.
- This could be attributed to the physical discrepancy between positions.
- Outside linemen are prototypically built to get to the QB faster and more often.
- Still, it is more valuable to be able to pressure from the outside and the frequency of pressure amplifies the value discrepancy.



How much does pressure affect an offense's ability to complete passes?

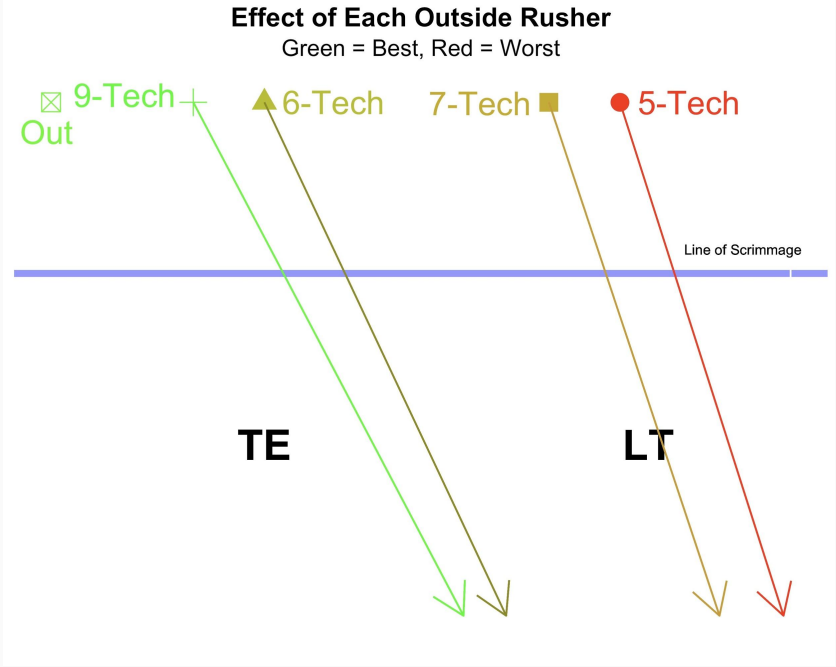
- We observe a similar trend when analyzing the effect of pressures on whether or not a pass will be completed through a logistic model.

- Among the three general groupings, outside linemen produce the greatest and most statistically significant effect on whether or not a pass will be completed.

```
Call:
glm(formula = as.numeric(Completion) ~ PrsByMiddle + PrsByInside +
    PrsByOutside + NoPrs, family = "binomial", data = passes)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5415  -0.9961   0.8527   0.8527   1.4312

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.44275    0.01893 -23.387  < 2e-16 ***
PrsByMiddle  -0.08190    0.10995  -0.745  0.45633
PrsByInside  -0.12450    0.07281  -1.710  0.08730 .
PrsByOutside -0.13652    0.04924  -2.772  0.00556 **
NoPrs         1.26724    0.02316  54.721  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Which technique within the outside group creates the most effective pressure?

- On average, the 9-Tech produces the greatest negative effect on EPA and Completion Percentage following the same process as the prior slides.



Pass Rushing Summary

- Given the defensive line's greatest way to affect a pass play is to get to the QB, pressures are the best stat to use to evaluate how each grouping affects passing plays.

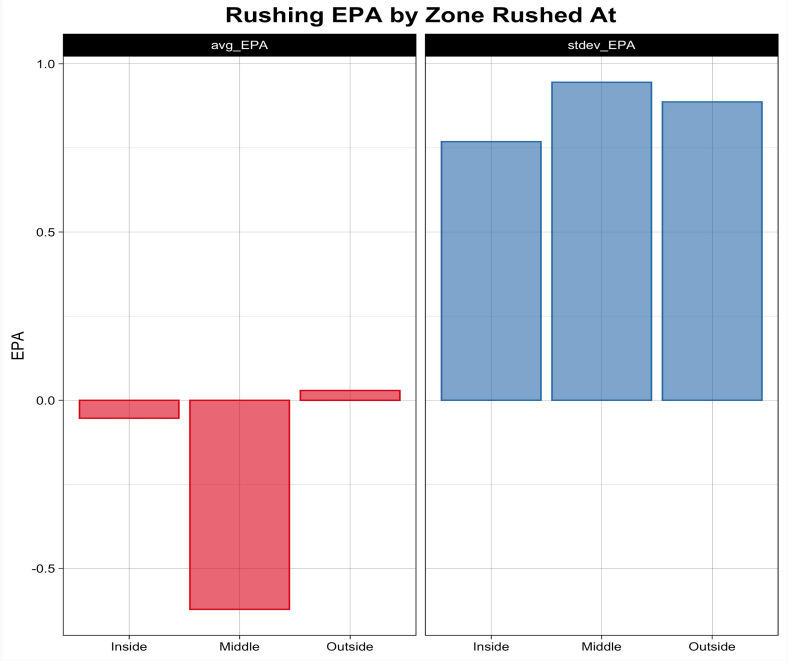
- Outside linemen affect passing plays through pressure more significantly and more often than the other groupings on the line.

- 9-Tech is a common positioning that is of the most advantageous for the defense to utilize against the pass.

What about against the run?

- We ought to make one distinguishment when analyzing runs plays: whether or not the designed gap was used.
  - This gives us a general proxy for whether or not the play went as the offense planned (this is an assumption).
  - If a particular position is 'better' at stopping an offense, even when the offense executes its designed play, that position should be considered more valuable.
- The data was cleaned so that we could identify three things:
  - What generalized area the rush went to.
  - If the rusher ran toward a defensive lineman or not.
  - Run play analysis was grouped so that we'd analyze the play to see the effect of the player only when the ball was ran at them. Meaning, a 0-Tech will not be held liable for a pitch play.

On executed rushes ran at each general zone, where do offenses have the most success?



- Rushes to the middle result in very low EPAs, on average, with more variance than runs to the outside and inside gaps.
- Executed outside runs provide the highest return with moderate variance.
    - This is likely due to the plethora of open space once you're able to get a running back to the edge.

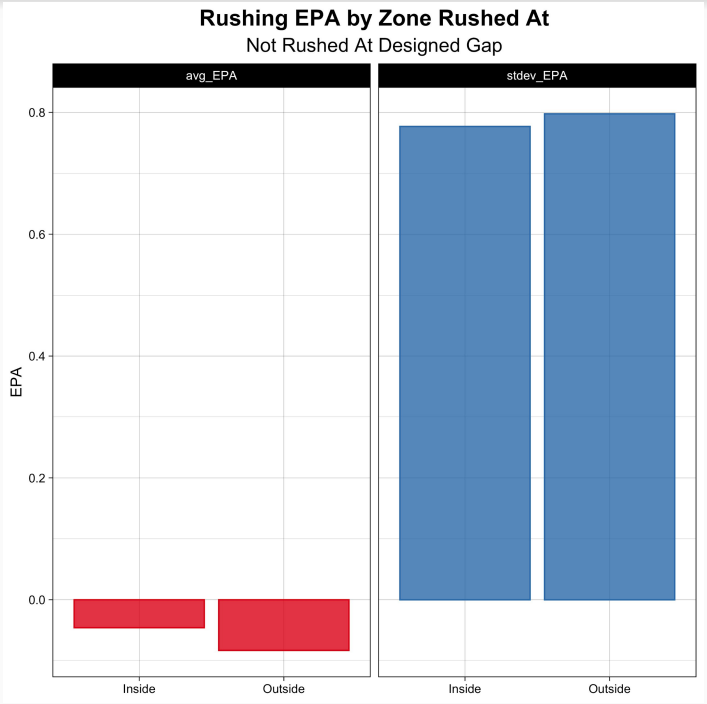
Regression shows us that runs to the outside are good for offenses, so defenses need to set the edge.

- The regression output shows that when the offense executes an outside run, it's worth a positive EPA to a statistically significant level.
- Given the high negative impact on EPA from middle rushes, we assume this is not an advantageous place for offenses to run given the dense nature of this area. With rushes to the middle and inside corresponding to a negative EPA, a defense should want opposing offenses to run to these areas.
- If we want to stop the run, controlling the edge looks like the most important area. If we don't set the edge, the run matters.

```
Call:
lm(formula = EPA ~ RanAtInside + RanAtOutside + RanAtMiddle,
    data = rushesExecuted)

Residuals:
    Min      1Q  Median      3Q     Max
-7.3509 -0.4191 -0.0941  0.4501  5.5146

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.037876   0.006395  -5.922 3.22e-09 ***
RanAtInside  -0.015558   0.027309  -0.570   0.5689
RanAtOutside  0.066821   0.018873   3.540   0.0004 ***
RanAtMiddle  -0.583498   0.048056 -12.142  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8769 on 22667 degrees of freedom
Multiple R-squared:  0.007202,  Adjusted R-squared:  0.007071
F-statistic: 54.81 on 3 and 22667 DF,  p-value: < 2.2e-16
```

Non-Executed Run Plays

- On non-executed rush plays we observe similar trends, but with less statistical confidence.
- Plays that aren't run to the designed gap are in the minority and exhibit more random behavior.
- Rushes to the middle were omitted due to a low sample of plays.
- We can continue with the selection of the outside section of the defensive line as our most important area.



Rush Defense Summary

- You'll want to be strongest where your opposition has the most potential to hurt you.
- In this case, runs to the outside are most advantageous for opposing offenses. Given this, having a strong presence on the edge in the run game appears most vital.

1. What position is most valuable?

Based on this research, I believe it's fair to conclude EDGE defenders are the most important defensive line positions. I am re-classifying the defensive line down into the following positions: Nose Tackle (0-Tech), Defensive Tackle (1-Tech through 3-Tech), Defensive End (4i through 5-Tech), and EDGE (any defender at 7-Tech or beyond). EDGE players can play both standing up and with their hand in the dirt, which allows for versatility in the run and pass-rush game. In each analysis we observed how the linemen farther from center had one of two things:

1. A greater history of affecting the play's outcome (completion or EPA) via the pass-rush, and
2. The need to seal the weakest part of run plays (setting the edge)

The run and pass were not created equal. The data we were provided reaffirms conventional analytical wisdom that passes are more valuable than runs when using EPA as your standard. Therefore, being able to affect the passer often is the most important thing a defensive line can do. EDGE defenders have the highest pressure rates. Add in that when EDGE defenders are the ones creating pressure, the opposing offense's EPA is expected to worsen to a statistically significant extent in comparison to defenders closer to center.

1. (cont.)

This analysis caused a shift in how I classified defensive line positions. I re-grouped the 5-Tech in with techniques that were subjectively placed on the interior originally. This is a change from my definitions from the beginning, but it became important to specify more directly the positions for the end result.

2. How does the distribution of talent across positions affect value?

The distinction between value/importance and talent is a crucial one. This project's first question is about measuring importance. Through the uneven nature of passing and running, we know being able to affect the quarterback is vital. Outside linemen tend to produce better results when they're the ones providing pressure. We also learned that the most dangerous run zone for a defense is the outside. This creates an inherent need for a strength on the outside. However, this should not be conflated with a statement about the talent of these positions.

Each defensive line position requires a different skillset. There is overlap between techniques played, but the distribution of talent is unique to each position. For instance, a nose tackle is at a disadvantage to produce the most value given what he's asked to do at that position and at his size. Being able to get to the quarterback quickly is better suited to a player who is 265 pounds opposed to a player who is 320 pounds.

This could amplify the importance of interior lineman in the run game, but more space on the edges results in more room to run for running backs. This requires the most versatile players to be on the outside. These players are slimmer in physique and faster. The game is better suited for their skillset to obtain higher value, opposed to one position being purely more talented than the other.

3. Exceptions

One of the greatest limitations of this form of league-wide analysis is its lack of specificity. Meaning, if Quenton Nelson is the Left Guard you're facing you can likely expect better results from rushes to the B (Inside) Gap. Being positioned strongly on the outside is the best on average, but when an offense is strongest at Center and Guard, your commitment to the outside won't pay as great of dividends.

In terms of realistic personnel, it is supremely difficult to find an EDGE lineman who can rush the passer at a high level and be a force in the run game consistently. With more space on the outside you need speed, but that speed comes at the cost of strength. When there's a 320 pound pulling guard it is difficult for your smaller EDGE lineman to get off that block.

Assumptions

- My general positional assignments encapsulate the most common gap assignments of the defensive line positions.
    - i.e. a player labelled as "Outside" for TechniqueName was assigned to the outside grouping which would significantly alter the results of the regression summaries and visualizations.
    - If a player labelled "Outside" is not to be considered a true defensive lineman, this would go against how I interpreted the positions and the definition of "Off-Ball".
- All assumptions related to Linear Regression (i.e. Linearity, Constant Variance, Normality).
- The use of indicator variables on run plays does subject us to bias. Because I chose to not hold a Nose Tackle liable for pitch plays this could results in not fully accounting for exceptional contributions on some plays.

Future Analysis

This data, in combination with defensive alignment data, yields very valuable information on what formations are better equipped to limit different offensive looks. In the current dataset, formation would need to be inferred from positioning of the line. If provided with different formations we would be better able to understand the strengths and weaknesses of different variations of 4-3, 3-4, and Nickel packages, for instance.

This would better show how many men are needed in the box to properly balance the asymmetric risk-reward relationship between stopping the pass versus stopping the run.

Supplements on Methodology

I didn't see an application for a very fancy machine learning model. Given this was mostly geared to study the effect different positions have on outcomes, regression was a simple way to test averages using a variety of indicator and categorical variables to link several parts of the data together.

For instance, several case_when and if_else statements were needed to derive when a run was targeted at a player given there are two sides (left and right) to the formation. These were the basis for building all features that were used as predictors in the models.

Thank you!

I appreciate you taking the time to read this analysis!

Feel free to reach out via email or connect with me on LinkedIn.

Email: calvinsmith625@gmail.com